Dynamic Hybrid Forecasting Models for Drug Consumption Prediction in Hospital Pharmacies

Yuxin Fan* and Siye Wu†
*School of Engineering and Applied Science, University of Pennsylvania, Canada, Toronto
Email: yuxinfan@alumni.upenn.edu
†Simon Business School, University of Rochester, Canada, Toronto
Email: april.siyewu@hotmail.com

***Abstract*—Accurate and efficient drug consumption forecasting is crucial for hospital pharmacies to avoid overstocking, minimize wastage, and ensure continuous patient care. This study proposes a hybrid forecasting framework integrating XGBoost, Prophet, and SARIMAX to improve monthly consumption predictions at the drug-manufacturer level. Through rolling-window forecasting and advanced feature engineering, the proposed approach addresses challenges such as seasonality, trend shifts, and sparse data. Experimental results demonstrate significant improvements in prediction accuracy and robustness across diverse drug consumption scenarios.**

***Index Terms*—Drug Consumption Forecasting, Hospital Pharmacies, Forecasting Models, XGBoost, Prophet, SARIMAX**

I. Introduction

Drug consumption forecasting plays a critical role in hospital pharmacy inventory management. Accurate drug consumption predictions enable hospital pharmacies to ensure drug availability while minimizing costs associated with overstocking or stockouts. As highlighted by Koala et al. [1], forecasting drug consumption is particularly challenging due to numerous influencing factors, such as sociodemographic characteristics, morbidity patterns, drug price index, and seasonal factors like disease outbreaks or policy changes, resulting in highly complex and dynamic consumption patterns. Therefore, the purpose of our study is to integrate a hybrid framework combining machine learning and statistical models to enhance drug consumption forecasting for hospital pharmacies.

In previous studies, various forecasting techniques have been proposed to address one or more of the above challenges. For instance, Taylor and Letham [2] introduced Prophet, a scalable forecasting method designed for large-scale applications, which excels at capturing general trends and seasonality through changepoint detection. However, Prophet often requires careful tuning to prevent overfitting to localized patterns and has limited capability in incorporating external variables, which are critical in drug consumption forecasting due to the influence of factors such as pricing and policies. Extreme Gradient Boosting (XGBoost), proposed by Chen and Guestrin [3], has demonstrated remarkable success in applications such as sales prediction and customer behavior modeling, highlighting its flexibility across diverse domains. However, XGBoost struggles with sequential dependencies in time-series data, making it less effective in forecasting tasks where temporal order plays a critical role, such as tracking drug consumption patterns over time.

The above limitations highlight a common challenge of single-model approaches, which often struggle to generalize across datasets exhibiting varying seasonality, non-linear dependencies, and sparsity. Therefore, some hybrid model frameworks have been introduced to provide more robust solutions. Comparative studies like those by Meng et al. [4] highlight that while Long Short-Term Memory (LSTM) demonstrates superior accuracy in drug sales forecasting, due to its ability to capture long-term dependencies and non-linear patterns, Prophet remains an effective tool for handling time series data with seasonality and holiday effects. However, both models exhibit limitations, such as LSTM's higher computational requirements and Prophet's sensitivity to trend and seasonality assumptions, which may hinder their adaptability in scenarios where drug consumption is influenced by dynamic external factors like policy changes or sudden disease outbreaks. Xu et al. [5] propose a hybrid approach that combines a linear regression (LR) model, specifically the auto-regression (AR) or auto-regressive integrated moving average (ARIMA) model, and a deep belief network (DBN) to improve time series forecasting by capturing both linear and nonlinear patterns. While this method demonstrates high accuracy in general time series prediction, its reliance on computationally intensive DBN training and limited consideration of irregular external factors, may pose challenges in hospital pharmacy drug consumption forecasting. Siddiqui et al. [6] propose a hybrid forecasting model, ARIMA-Holt's Winter (ARHOW), which combines the strengths of the ARIMA model and the Holt-Winters method to enhance forecasting accuracy. This study, specifically applied to the pharmaceutical industry, demonstrates that this hybrid model involves high demand volatility and complex seasonal patterns. However, the reliance on fixed model architectures and pre-determined parameter settings may limit its adaptability in highly dynamic environments. Rathipriya et al. [7] develop a hybrid demand forecasting model using both shallow neural networks like Radial Basis Function Neural Network (RBF_NN) and Generalized Regression Neural Network (GR_NN), and deep learning models like LSTM and Stacked LSTM to predict pharmaceutical sales. Their study highlights the effectiveness of shallow neural networks in handling smaller datasets and mitigating noise,

while deep learning models like LSTM excel at capturing non-linear and temporal patterns in larger datasets. However, the computational complexity of deep learning models and their dependency on large datasets may pose challenges in hospital pharmacy drug consumption forecasting, where data sparsity and rapid external changes often occur.

In this study, to address the limitations of existing models, such as adjusting predictions for non-stationary data, capturing shifting trends dynamically, and accounting for diverse drug-manufacturer combinations, our proposed framework introduces a new hybrid model called DynamicXSP integrating XGBoost, Prophet, and Seasonal AutoRegressive Integrated Moving Average with eXogenous variables (SARIMAX) with rolling-window forecasting, advanced feature engineering and grid search. The dataset used for this study includes monthly drug consumption records collected from hospital pharmacies over a period of nearly seven years, including a wide variety of drugs and manufacturers. Models in DynamicXSP like Prophet excel at capturing stable long-term trends and seasonal patterns, making them suitable for drugs with high and regular demand. On the other hand, SARIMAX is better equipped to handle datasets influenced by external variables or strong short-term dynamics, as it integrates advanced feature engineering and external factor modeling. For scenarios with complex non-linear relationships or sparse observations, XGBoost proves advantageous due to its flexibility in handling feature interactions and non-linear dependencies. Therefore, DynamicXSP enriches the data with lag features and trend indicators, capturing both short-term dependencies and long-term dynamics. This synergy provides a robust and versatile solution to the unique challenges of drug consumption forecasting, making the framework applicable across various drug types and manufacturers.

The remainder of this paper is organized as follows: Section II details the methodology, highlighting the advantages of the proposed framework. Section III presents experimental and results, including data preprocessing and data analysis. Finally, Section IV concludes the study and outlines future research directions.

II. Methodology

A. DynamicXSP Model Framework

The DynamicXSP framework integrates SARIMAX, XGBoost and Prophet to improve drug consumption forecasting by leveraging the strengths of each model. Given the heterogeneous nature of drug-manufacturer combinations, where some exhibit stable trends while others fluctuate due to supply chain disruptions or policy changes, a dynamic approach is necessary.

To optimize predictive accuracy, the framework evaluates multiple models for each combination using performance metrics such as R^2 and symmetric mean absolute percentage error (SMAPE). The best-performing model is selected to ensure adaptability to varying consumption patterns.

DynamicXSP follows a structured process:

- Feature engineering constructs lag-based variables, rolling statistics, and external factor indicators to enhance predictive modeling.
- Model training applies SARIMAX, XGBoost and Prophet with rolling-window forecasting to ensure adaptability to recent trends.
- Model selection identifies the most effective model for each drug-manufacturer pair based on validation performance.

By dynamically assigning models based on data characteristics and performance metrics, the framework enhances predictive accuracy and robustness in hospital pharmacy inventory management.

B. Model Design and Roles

1) DynamicXSP-S (SARIMAX) Model: SARIMAX extends the traditional ARIMA framework by incorporating external (exogenous) variables, making it particularly effective for forecasting drug consumption, where demand is influenced by both historical trends and external factors such as manufacturer-specific supply patterns and hospital inventory policies. The model is defined as:

$$y_t = \phi(B)\theta(B)^{-1}\left(c + \mathbf{X}_t\beta + \epsilon_t\right), \tag{1}$$

where y_t represents the weekly reduction in drug inventory, $\mathbf{X}_t$ denotes the set of exogenous features designed to capture relevant consumption patterns, and ϵ_t is a white noise error term.

To improve predictive performance, our DynamicXSP-S model incorporates domain-specific feature engineering tailored to hospital pharmacy drug consumption data. These features capture both short-term fluctuations and long-term trends, ensuring robustness in dynamic demand environments. The key engineered features include:

- Lagged values: Capture delayed effects of past consumption, particularly useful for modeling hospital restocking cycles and manufacturer supply constraints:

$$\text{drug\_consume\_lag}_k = y_{t-k}, \quad k \in \{1, 3, 6, 12\}. \tag{2}$$

  These lag selections align with common drug procurement intervals observed in historical data.
- Rolling statistics: Represent short-term demand fluctuations and variability, helping the model account for unexpected surges in drug usage due to seasonal illnesses or policy changes:

$$\text{consume\_rolling\_mean}_k = \frac{1}{k}\sum_{i=1}^{k} y_{t-i}, \tag{3}$$

$$\text{consume\_rolling\_std}_k = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(y_{t-i} - \text{mean})^2}. \tag{4}$$

  Unlike standard moving averages, these rolling features dynamically adjust to different drug types, ensuring adaptability across various manufacturers.
- Seasonality encoding: Drug consumption often follows predictable monthly cycles due to hospital procurement

schedules. To effectively model this, we apply trigonometric transformations:

$$\text{month\_sin} = \sin\left(\frac{2\pi \cdot \text{month}}{12}\right), \tag{5}$$

$$\text{month\_cos} = \cos\left(\frac{2\pi \cdot \text{month}}{12}\right). \tag{6}$$

This formulation allows SARIMAX to capture periodic fluctuations smoothly without requiring manual seasonal differencing.

- Exponentially Weighted Moving Average (EWMA): Provides a memory-adjusted smoothing technique that prioritizes recent observations, making the model more responsive to sudden shifts in drug demand:

$$\text{EWMA}_\alpha = \alpha y_t + (1-\alpha) \cdot \text{EWMA}_{\alpha,t-1}. \tag{7}$$

The smoothing factor α is adjusted based on drug-specific volatility (in our model, we use 3), ensuring that short-term demand spikes are reflected in the model while filtering out noise.

- Percentage change and trend strength: Designed to measure relative variations and stability in consumption patterns, providing key signals for adaptive inventory planning:

$$\text{pct\_change\_consume}_k = \frac{y_t - y_{t-k}}{y_{t-k}}, \tag{8}$$

$$\text{comsume\_trend\_strength} = \frac{1}{k}\sum_{i=1}^{k}|y_{t-i} - y_{t-i-1}|. \tag{9}$$

The trend strength metric allows the model to differentiate between regular seasonal variations and abrupt changes due to supply chain disruptions.

These engineered features were selected based on statistical analysis of hospital pharmacy consumption data, ensuring that the SARIMAX model captures both recurrent and unexpected demand shifts. By integrating rolling-window forecasting and dynamically adjusting feature selection per drug-manufacturer pair, our approach enhances predictive accuracy and robustness in real-world hospital inventory management scenarios.

2) DynamicXSP-X (XGBoost) Model: XGBoost is a gradient boosting framework that constructs an ensemble of decision trees to predict the target variable y_t. It is particularly effective for time-series forecasting involving non-linear relationships and sparse data, making it well-suited for drug consumption prediction, where demand varies across manufacturers and fluctuates due to external factors. The model predicts y_t through an additive function:

$$\hat{y}_t = F(x_t) = \sum_{k=1}^{K} f_k(x_t), \quad f_k \in \mathcal{F}, \tag{10}$$

where $\hat{y}_t$ is the predicted value, x_t represents input features, f_k denotes the k-th decision tree, and $\mathcal{F}$ is the space of decision trees.

To effectively capture temporal dependencies, our model utilizes lagged values ($y_{t-1}, y_{t-2}, \ldots$), rolling statistics (e.g., moving averages and standard deviations over 3, 6, and 12 periods), and consumption trend indicators. These features help the model recognize historical patterns in drug usage and detect shifts in demand dynamics, which is essential given the variability across different drug manufacturers. Additionally, interaction terms between lagged values and rolling statistics further enhance the model's ability to capture complex dependencies.

Hyperparameters such as the number of estimators, learning rate, tree depth, and sampling ratios are tuned using grid search to achieve a balance between predictive accuracy and model complexity. As part of the hybrid forecasting framework, our DynamicXSP-X benefits from dynamic updates to its training data, allowing it to remain responsive to evolving demand patterns.

By capturing non-linear interactions and irregular consumption trends, DynamicXSP-X serves as a powerful complement to statistical models like DynamicXSP-S. Its ability to learn complex patterns from historical data ensures that it remains a robust and adaptive component of the forecasting system.

3) DynamicXSP-P (Prophet) Model: Prophet is a time-series forecasting model, designed to decompose data into trend, seasonality, and external event components. In our model, we optimized Prophet to DynamicXSP-P to make it more effective for handling missing values, outliers, and irregular consumption patterns for drug demand forecasting, where sudden fluctuations and manufacturer-specific trends are common. The model predicts the target variable y_t as:

$$y_t = g(t) + s(t) + h(t) + \epsilon_t, \tag{11}$$

where $g(t)$ represents the long-term trend, $s(t)$ captures seasonal patterns using Fourier series, $h(t)$ accounts for external disruptions (such as regulatory policy changes or supply chain delays), and ϵ_t is the white noise error term. This decomposition enhances interpretability while allowing the model to dynamically adjust to varying consumption behaviors.

Key hyperparameters are optimized to ensure the model adapts effectively to changes in drug demand. The tuning process focuses on:

- *seasonality_mode*: Determines whether seasonal effects are additive or multiplicative, depending on consumption volatility.
- *changepoint_prior_scale*: Controls the sensitivity to abrupt trend shifts, crucial for modeling demand surges or supply disruptions.
- *seasonality_prior_scale*: Adjusts the weight of seasonal components to ensure a balance between smooth long-term trends and short-term variations.

A higher *changepoint_prior_scale* allows the model to react more quickly to structural changes, such as increased demand due to new hospital procurement policies, while lower values favor smoother trend transitions.

DynamicXSP-P benefits from continuous data updates, enabling it to remain responsive to evolving drug consumption trends while preventing overfitting to outdated patterns. By leveraging decomposition-based forecasting and hyperparameter tuning, Prophet provides an interpretable and adaptive prediction mechanism that complements DynamicXSP-S and

DynamicXSP-X within the hybrid framework. Its ability to capture trend shifts and seasonal dependencies makes it a crucial component for enhancing forecasting accuracy across various drug manufacturers.

C. Dynamic Rolling-Window Forecasting

Time-series data in drug consumption forecasting often exhibit non-stationarity, where trends, seasonality, and noise evolve over time. Static forecasting approaches that rely on fixed historical data may struggle to capture these dynamics, leading to suboptimal performance. To address this, a dynamic rolling-window forecasting strategy is employed, enabling models to prioritize recent information and adapt to structural changes in demand.

At each time step t, the training dataset is updated to include the most recent observations while discarding older data beyond the defined window size (W). Formally, the training dataset at t is defined as:

$$\mathcal{D}_t = \{(y_\tau, \mathbf{X}_\tau) \mid \tau \in [t - W, t - 1]\},$$

where $\mathcal{D}_t$ is the training data, y_τ denotes the target variable (drug consumption), and $\mathbf{X}_\tau$ represents the corresponding feature vectors. After training on $\mathcal{D}_t$, predictions are generated for the next time step $(t + 1)$.

The rolling-window approach effectively captures short-term dynamics by prioritizing recent patterns while mitigating the influence of outdated data. This is particularly beneficial for handling structural changes such as sudden shifts in demand due to hospital procurement cycles, manufacturer-specific supply fluctuations, and regulatory interventions. The choice of window size (W) is critical: larger windows incorporate long-term trends, while smaller windows emphasize recent changes. This study evaluates multiple window sizes and selects the optimal configuration based on metrics such as root mean squared error (RMSE) and symmetric mean absolute percentage error (SMAPE).

Integrating rolling-window forecasting into the hybrid framework enhances adaptability:

- **DynamicXSP-S**: Dynamically recalibrates coefficients to model short-term dependencies and external influences in drug consumption.
- **DynamicXSP-X**: Refines decision trees with updated feature interactions, capturing evolving non-linear relationships in manufacturer-specific demand shifts.
- **DynamicXSP-P**: Updates its trend decomposition to align with the latest data, ensuring better adaptability to seasonal and structural variations.

This iterative strategy ensures that the models remain responsive and resilient to changes, achieving superior forecasting performance across diverse drug consumption scenarios by continuously adapting to the most recent data trends.

III. Experiments

A. Dataset

The dataset used in this study consists of monthly drug consumption records collected from various hospital pharmacies between January 1, 2018, and September 1, 2024. The data include a diverse range of drugs across multiple categories and manufacturers, ensuring a comprehensive representation of pharmaceutical demand. Each record contains the drug name, manufacturer, monthly consumption values, inventory levels, and associated features such as proportions and trends. This dataset provides a robust foundation for developing and testing the proposed forecasting framework.

B. Data Pre-processing

A comprehensive preprocessing pipeline was applied to ensure data quality and consistency, while focusing on key steps relevant to our framework. Missing values in critical features were interpolated where possible, and records with excessive missing data were excluded to maintain dataset integrity.

[name/specify the method of outliers] Outliers in consumption data were addressed using a rolling-window approach. For each drug-manufacturer combination, rolling statistics such as mean and standard deviation were computed over a seven-month window. Values exceeding three standard deviations from the mean or falling outside the 5th and 95th percentiles were flagged as anomalies and adjusted to boundary values to preserve the time series' continuity.

To ensure the dataset included only high-quality samples for modeling, groups with fewer than six months of non-zero consumption data or those exhibiting extreme sparsity were excluded. Temporal autocorrelation of consumption data was assessed, and groups failing to exhibit sufficient autocorrelation were removed. Additional filtering criteria included variance thresholds and limits on skewness to avoid heavily imbalanced target distributions.

Feature engineering was performed to enhance the dataset's predictive power. Derived features such as lagged consumption values (e.g., previous month's consumption), rolling statistics (e.g., mean, variance, and percent changes), and seasonality indicators encoded with trigonometric functions were created to capture temporal dependencies and periodic trends. These steps ensured that the final dataset was robust, informative, and well-suited for downstream predictive modeling tasks.

C. Data Analysis

To improve the effectiveness of forecasting models on drug consumptions, we implemented DynamicXSP model framework. Each model in Dynameic was applied to forecast sales data for a diverse set of drugs and manufacturers. The dataset consisted of weekly time series data spanning seven years, capturing trends, seasonality, and irregular fluctuations in drug consumptions. The selection of the optimal model in DynamicXSP for each drug-manufacturer pair was based on achieving the highest R^2 and minimizing SMAPE.

D. Model Results

1) Representative Drug Cases:

a) DynamicXSP-X (XGBoost) Model:

- **Drug:** Mycophenolate Sodium Enteric Tablets
  - **Manufacturer:** Novartis Switzerland
  - **Metrics:** $R^2 = 0.8154$, SMAPE = 25.40

XGBoost showed exceptional performance for this drug by effectively capturing the complex non-linear patterns and sudden changes in demand, particularly evident in the volatile periods of 2023-2024.

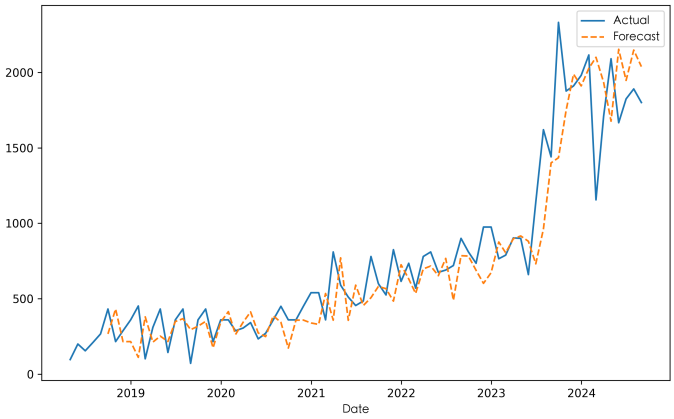

Fig. 1. XGBoost Prediction for Mycophenolate Sodium Enteric Tablets by Novartis.

- **Drug:** Flurbiprofen Gel Patch
  - **Manufacturer:** Jingtaide
  - **Metrics:** $R^2 = 0.7902$, SMAPE = 21.14

XGBoost effectively handled the increasing trend and high volatility in sales, demonstrating its strength in capturing complex patterns and sudden market changes.

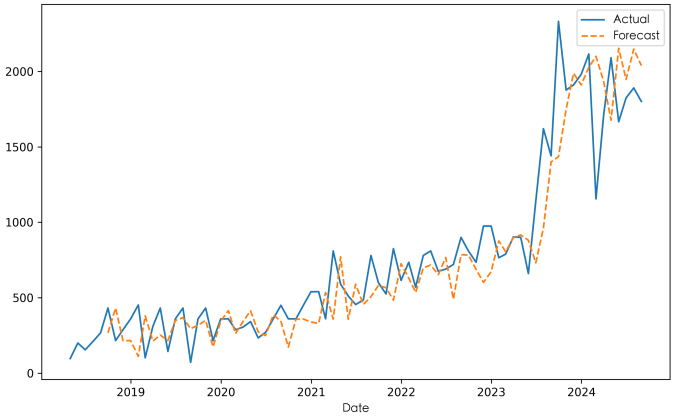

Fig. 2. XGBoost Prediction for Flurbiprofen Gel Patch by Jingtaide.

b) DynamicXSP-P (Prophet) Model:

- **Drug:** Compound Phellodendron Liquid
  - **Manufacturer:** Lu Hanfang
  - **Metrics:** $R^2 = 0.7898$, SMAPE = 22.63

Prophet excelled in modeling this drug's sales pattern due to its robust handling of trend changes and seasonal variations, particularly evident in the steady growth pattern.

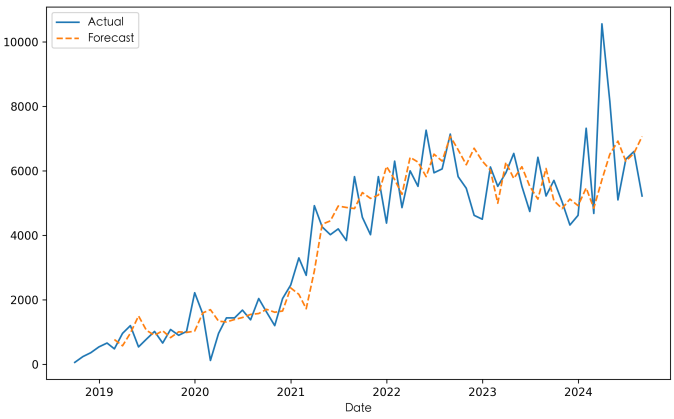

Fig. 3. Prophet Prediction for Compound Phellodendron Liquid by Lu Hanfang.

- **Drug:** Shenshuaining Tablets
  - **Manufacturer:** Shanhaiguan Pharmaceutical
  - **Metrics:** $R^2 = 0.7348$, SMAPE = 21.86

Prophet's capability to handle multiple seasonality levels and trend changes made it the optimal choice for this drug's complex sales pattern.

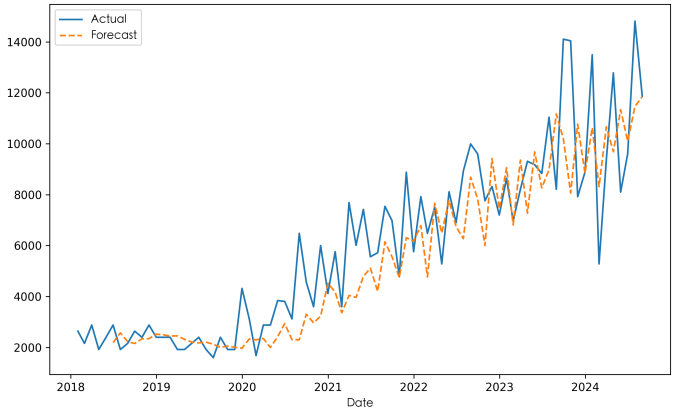

Fig. 4. Prophet Prediction for Shenshuaining Tablets by Shanhaiguan Pharmaceutical.

c) DynamicXSP-S (SARIMAX) Model:

- **Drug:** Peritoneal Dialysis Solution [Lactate]
  - **Manufacturer:** Huaren
  - **Metrics:** $R^2 = 0.8109$, SMAPE = 32.49

SARIMAX demonstrated superior performance for this drug due to its ability to capture both the seasonal patterns and the gradual decline in demand shown in the time series. The model effectively handled the relatively stable periodic fluctuations until the sharp decline in 2023.

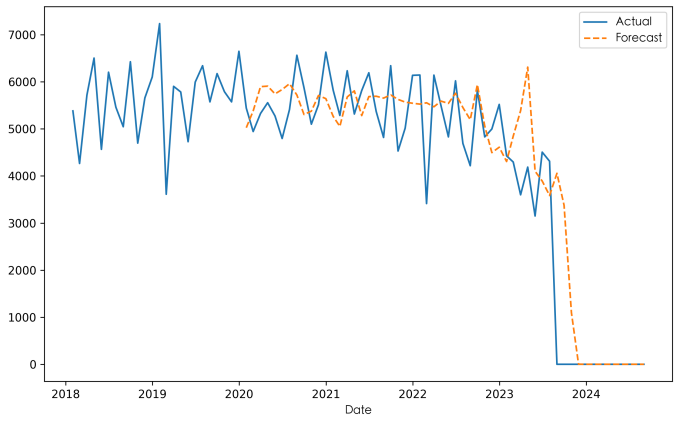

Fig. 5. SARIMAX Prediction for Peritoneal Dialysis Solution by Huaren.

- **Drug:** Vitamin B1 Tablets
  - **Manufacturer:** Xinyi Huanghe
  - **Metrics:** $R^2 = 0.7345$, SMAPE = 35.37

  The SARIMAX model was effective for this drug due to its ability to handle the external influences and volatility in the sales data.

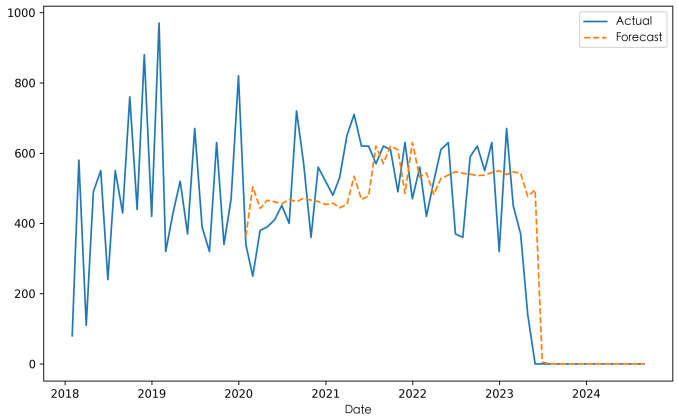

Fig. 6. SARIMAX Prediction for Vitamin B1 Tablets by Xinyi Huanghe.

2) Overall Results: The results highlighted that no single model universally outperformed the others across all cases, underscoring the importance of tailoring the choice of forecasting models to specific data characteristics. SARIMAX demonstrated superior capability in capturing seasonality influenced by external variables, such as pricing trends and policy changes. XGBoost excelled in modeling complex, non-linear relationships within the data, making it particularly effective for patterns with high variability. Prophet performed well in cases with dominant seasonality and long-term trends, such as periodic spikes in drug demand. These findings proves the necessary of DynamicXSP by leveraging the complementary strengths of each single model.

The visualization analysis reveals several key findings:

Long-term Trend Capture:

- All models effectively captured the underlying growth trends, with Prophet showing particular strength in long-term pattern recognition.
- SARIMAX demonstrated superior performance in capturing gradual trend changes, as evidenced in the insulin device forecasts.

Seasonal Pattern Recognition:

- XGBoost effectively captured complex seasonal patterns in chronic medication usage.
- SARIMAX showed strong performance in regular seasonal variations, particularly in established products.

Growth Pattern Adaptation:

- Prophet demonstrated superior capability in adapting to emerging growth patterns.
- XGBoost showed strong performance in capturing non-linear growth relationships.

IV. Conclusion and Future Work

A. Conclusion

The results of this study underscore the effectiveness and adaptability of the DynamicXSP framework in addressing the challenges of pharmaceutical demand forecasting. By integrating the complementary strengths of SARIMAX, Prophet, and XGBoost, along with advanced techniques such as rolling-window forecasting, grid search and feature engineering, DynamicXSP provides a robust solution for predicting drug consumption across diverse drug-manufacturer combinations.

This framework's tailored approach allows the selection and integration of models based on the specific characteristics of each dataset, enabling it to overcome limitations inherent in single-model approaches. The performance evaluation highlights the key contributions of each model within the framework:

- **DynamicXSP-X (XGBoost) Model**: Demonstrated superior handling of nonlinear relationships, particularly excelling in short-term predictions, where it outperformed SARIMAX and Prophet in terms of RMSE and SMAPE.
- **DynamicXSP-S (SARIMAX) Model**: Excelled in datasets dominated by seasonal trends and where external (exogenous) variables, such as pricing or policy changes, played a significant role.
- **DynamicXSP-P (Prophet) Model**: Effectively captured long-term seasonality and trends, proving valuable for datasets with prominent periodic patterns, though it faced challenges with sparsity.

The DynamicXSP framework's ability to dynamically adapt to various data characteristics and leverage the complementary strengths of multiple models highlights its practicality and scalability for hospital pharmacy forecasting. This hybrid approach not only improves prediction accuracy and robustness but also sets a foundation for broader applications in healthcare demand forecasting.

B. Future Work

To further improve drug inventory predictions, the following directions are proposed for future research:

- **Hybrid Framework Expansion**: Extend the hybrid framework by integrating additional models, such as LightGBM, Transformer-based architectures, or ensemble

methods to capture both short-term fluctuations and long-term trends.

- **External Variable Enrichment**: Explore additional exogenous variables such as patient inflow, epidemic outbreak data, seasonal illnesses (e.g., flu seasons), or hospital-specific factors to improve prediction accuracy.
- **Automated Model Selection**: Implement automated hyperparameter tuning and model selection techniques (e.g., Bayesian optimization) to improve forecasting performance with minimal manual intervention.

By addressing these areas, the proposed framework can become a more robust and scalable solution for hospital pharmacy inventory management, ensuring operational efficiencies of hospital pharmacies.

V. References

References

[1] D. Koala, Z. Yahouni, G. Alpan, and Y. Frein, "Factors influencing drug consumption and prediction methods," in *CIGI-Qualita: Conférence Internationale Génie Industriel QUALITA*, Grenoble, France, 2021.

[2] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.

[3] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[4] J. Meng, Q. Zhang, and X. Li, "Comparative analysis of Prophet and LSTM models in drug sales forecasting," *Journal of Physics: Conference Series*, vol. 1910, no. 1, p. 012059, 2021.

[5] W. Xu, Y. Wang, and J. Zhao, "A hybrid modelling method for time series forecasting based on a linear regression model and deep learning," *Applied Intelligence*, vol. 49, no. 7, pp. 2875–2888, 2019.

[6] R. Siddiqui, A. Khan, and M. Ahmed, "A Hybrid Demand Forecasting Model for Greater Forecasting Accuracy: The Case of the Pharmaceutical Industry," *Supply Chain Forum: An International Journal*, vol. 22, no. 3, pp. 1–13, 2021.

[7] R. Rathipriya, M. Saranya, and K. Ramkumar, "Demand Forecasting Model for Time-Series Pharmaceutical Data Using Neural Networks," *Neural Computing and Applications*, vol. 35, pp. 1945–1957, 2022.

Appendix

The following pseudocode outlines the sample selection process used to ensure the quality and relevance of the dataset for modeling tasks:

Algorithm 1 Sample Selection Algorithm

Require: Cleaned data df, configuration thresholds config
Ensure: Filtered dataset $final\_df$

$final\_df \leftarrow \emptyset$
for each unique combination of drug name and manufacturer (d, m) in df **do**
  $group\_data \leftarrow$ subset of df for (d, m)
  if length of $group\_data <$ config.min_months **or** sum of consumption $= 0$ **then**
    Skip group ▷ Insufficient or sparse data
  end if
  $non\_zero\_ratio \leftarrow$ proportion of non-zero consumption in $group\_data$
  if $non\_zero\_ratio <$ config.sparsity_threshold **then**
    Skip group ▷ Data too sparse
  end if
  $acf\_values \leftarrow$ autocorrelation function of consumption in $group\_data$
  if $\max(acf\_values[1:]) <$ config.min_acf_threshold **then**
    Skip group ▷ Insufficient autocorrelation
  end if
  if no start date $\geq$ config.min_start_date in $group\_data$ **then**
    Skip group ▷ No recent data
  end if
  $variance \leftarrow$ variance of consumption in $group\_data$
  if $variance <$ config.min_variance_threshold **then**
    Skip group ▷ Variance too low
  end if
  $missing\_ratio \leftarrow$ maximum missing ratio for features in $group\_data$
  if $missing\_ratio >$ config.max_missing_ratio **then**
    Skip group ▷ Feature missing data too high
  end if
  $skewness \leftarrow$ skewness of consumption in $group\_data$
  if $|skewness| >$ config.max_skewness **then**
    Skip group ▷ Target variable too skewed
  end if
  $correlation \leftarrow$ correlation of consumption with lagged features in $group\_data$
  if $|correlation| <$ config.min_correlation **then**
    Skip group ▷ Insufficient correlation with features
  end if
  $final\_df \leftarrow final\_df \cup group\_data$
end for
return $final\_df$